

# POLYGENIC AND MULTIFACTORIAL DISEASE GENE ASSOCIATION IN MAN: Lessons from AIDS\*

---

Stephen J. O'Brien<sup>1</sup>, George W. Nelson<sup>2</sup>,  
Cheryl A. Winkler<sup>2</sup>, and Michael W. Smith<sup>2</sup>

<sup>1</sup>Laboratory of Genomic Diversity, National Cancer Institute, <sup>2</sup>Intramural Research Support Program, SAIC Frederick, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201;  
e-mail: obrien@ncifcrf.gov

**Key Words** AIDS, SNP, restriction gene, complex traits

■ **Abstract** In an age when the majority of monogenic human disease genes have been identified, a particular challenge for the coming generation of human geneticists will be resolving complex polygenic and multifactorial diseases. The tools of molecular and population genetic association have much potential as well as peril in uncovering small cryptic genetic effects in disease. We have used a candidate gene approach to identify eight distinct human loci with alleles that in different ways influence the outcome of exposure to HIV-1, the AIDS virus. The successes in these gene hunts have validated the approach and illustrate the strengths and limitations of association analysis in an actual case history. The integration of genetic associations, well-described clinical cohorts, extensive basic research on AIDS pathogenesis, and functional interpretation of gene connections to disease offers a formula for detecting such genes in complex human genetic phenotypes.

## INTRODUCTION

The identification of human genetic variants that influence the outcomes of HIV-1 exposure graphically illustrates the power of human genomic assessment on human disease. The new advances have drawn together the findings of several disciplines to identify genetic polymorphisms as important regulators of how quickly individuals develop AIDS, of which AIDS-defining conditions develop and, in certain cases, whether individuals exposed to HIV-1 become infected on exposure to the virus (34, 75, 84, 85). The findings have implications not only for the understanding of the process by which HIV replication and dissemination lead to the collapse

---

\*The US Government has the right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

of the CD4+ T-lymphocyte population, a hallmark of AIDS, but also for identifying new targets for therapy development.

The discoveries emerged from a 15-year-long assembly of patient materials that culminated in the late 1990s with the intersection of four rapidly developing biological fields: (a) the exquisite cell biology, virology, and x-ray crystallography that identified the chemokine receptors, CCR5 and CXCR4, as requisite coreceptors or entry portals for HIV-cell entry; (b) the longitudinal detailing of clinical information on thousands of HIV-infected patients monitored in epidemiological cohort studies; (c) the high throughput genotyping and polymorphism discovery methodologies developed by the human molecular genetics community; and (d) the development and implementation of computational algorithms designed to employ population genetic theory to detect gene associations in disease cohort populations.

The human genome consists of 25,000 to 50,000 genes, ranging in length from a few hundred to several million nucleotides, arranged in linear arrays along 24 unique human chromosomes. Today a worldwide scientific endeavor called the Human Genome Project is near completion of its 15-year goal: to produce a full-length sequence of the 3.2 billion base pairs that comprise the human genome (21, 22). Scheduled for draft completion this year, the effort has already revealed relatively common single nucleotide polymorphism (SNP) variants every 500 to 1000 nucleotides (64, 117). This translates into an estimated 1.5–2.0 million genetic differences between any two people. Relating these differences to human phenotypes offers the alluring prospect for a genetic understanding of the basis for human distinctions in appearance, talent, behavior, hereditary disease, inflammatory reaction, and host response to infectious diseases including HIV.

Although the sequence of the human gene vocabulary is in hand, we still have names for fewer than 8000 genes, 15–20% of the total. A few hundred SNPs have been implicated in human hereditary diseases. A common belief is that nearly all of the genetically simple monogenic diseases have been identified or soon will be. The challenge of the next generation will certainly be polygenic (or complex) disease phenotypes (i.e. those with multiple loci exerting small quantifiable effects on the phenotype) and multifactorial diseases, those which require an environmental stimulus or cofactor for gene influence. Multifactorial disease genes mediate infectious disease outcomes as well as susceptibility to chemical toxins, such as smoking and lung cancer. These phenotypes are not always suitable for pedigree analysis, since susceptibility/resistance phenotypes of unexposed individuals are not accessible. For example, if we inspect a number of gay men infected with HIV-1, their parents and offspring (which are few) are not generally exposed. For such diseases and other multifactorial phenotypes, population association analysis is employed to discover genetic effects. In such cases, distortion of population genetic equilibria [(i.e. allele, genotype frequency, Hardy Weinberg equilibrium (HWE), and linkage disequilibrium (LD)] become tools of gene mapping. In this

review, we describe the association methods our collaborators and we employed to implicate gene effects on HIV-1 exposed individuals. We discuss their strengths and limitations and highlight the lessons learned from a field test of the tools for mapping complex multifactorial alleles that influence the AIDS epidemic.

## CANDIDATE GENES FOR AIDS

Association analysis can involve either candidate genes or genome scans of anonymous DNA markers that track disease susceptibility genes by linkage disequilibrium. (We discuss genome scans later in this review.) Candidate genes are screened for common SNPs by several methods including single strand conformation polymorphism (SSCP) (40, 52, 90), denaturing high pressure liquid chromatography (52, 98, 113), direct sequencing, and *in silico* consulting (the inspection of internet databases for cumulative lists of discovered and mapped human SNPs) (36, 72, 108, 121). Useful SNPs can be discovered in coding regions, in introns, in untranslated regions of transcripts, in upstream regulatory or promoter regions, and even among flanking microsatellite or short tandem repeat (STR) loci (36, 72, 121). A vast literature of AIDS pathogenesis has suggested several hundred candidate genes including HIV-1 coreceptors, their ligands, cytokines and their receptors, transcription factors, immune response genes, and other factors that participate in HIV-1-mediated immune destruction. SNPs have been described in scores of these loci and nearly all are being tested or already have been for influence on HIV pathogenesis.

The approach taken by our group was to assemble large cohorts of patients at risk for HIV-1 infection and to create B-cell lines from each as a renewable store of DNA for population genetic association analysis. In all, we have worked with some 20 different AIDS cohorts, including homosexual men, IV drug users, and hemophiliac patients exposed to contaminated clotting factor lots before the introduction of the HIV blood test in 1984. Detailed descriptions of the AIDS cohorts have been published (12, 29, 41, 42, 45, 53, 60, 88, 116). In total, our laboratory has collected some 10,000 individuals who have formed the population basis for our search. Using these large at-risk populations, clinically monitored over the course of HIV-1 exposure, infection, and disease progression, we related the population distribution of allele, genotype, and haplotype combinations to differential disease outcomes. The large numbers of individuals studied, whether pooled in statistical analyses or considered separately as replicate cohort studies, contributed to confidence that the observed effects were real and relevant to the disease we studied.

AIDS is a complex and chronic disease triggered by initial infection with HIV-1, which leads to depletion of the CD4-T-lymphocyte cell population as a prelude to immune system collapse. Genetic influence on the epidemiologic heterogeneity of AIDS could be discovered at any stage. To keep the search focused we concentrated

**TABLE 1** Epidemiologic heterogeneity in AIDS upon which genetic hypotheses were inspected

1. HIV infection and transmission;
2. The rate of progression to full-blown AIDS after HIV infection;
3. The AIDS sequelae or AIDS-defining disease at diagnosis (i.e., Kaposi's sarcoma, *Pneumocystis carinii* pneumonia, lymphoma, opportunistic infection, neuropathology, etc.);
4. The immune response, cellular and humoral to HIV;
5. Efficacy and side effects of highly active antiretroviral therapy (HAART).

on the five discrete stages of AIDS based on observed epidemiologic patterns listed in Table 1. Our approach was to discover common SNP (or insertion/deletion) variants in candidate restriction genes and to test their influence, alone and interacting with each other, on the five steps in Table 1 using the HIV-1/AIDS cohort populations. The first AIDS restriction allele, *CCR5*- $\Delta$ 32, was discovered in 1996 as a recessive influence that blocked HIV-1 infection (24, 66, 96, 128). We now know that *CCR5*- $\Delta$ 32 also affects both the rate of disease progression and the incidence of AIDS-associated B-cell lymphoma in HIV-1-infected heterozygotes (24, 25, 66, 96, 128). Several additional AIDS restriction genes were subsequently discovered to affect different points of AIDS progression and HIV-1 infection (Table 2).

**TABLE 2** Identified genes that affect HIV-1 infection, progression to AIDS, and AIDS outcome

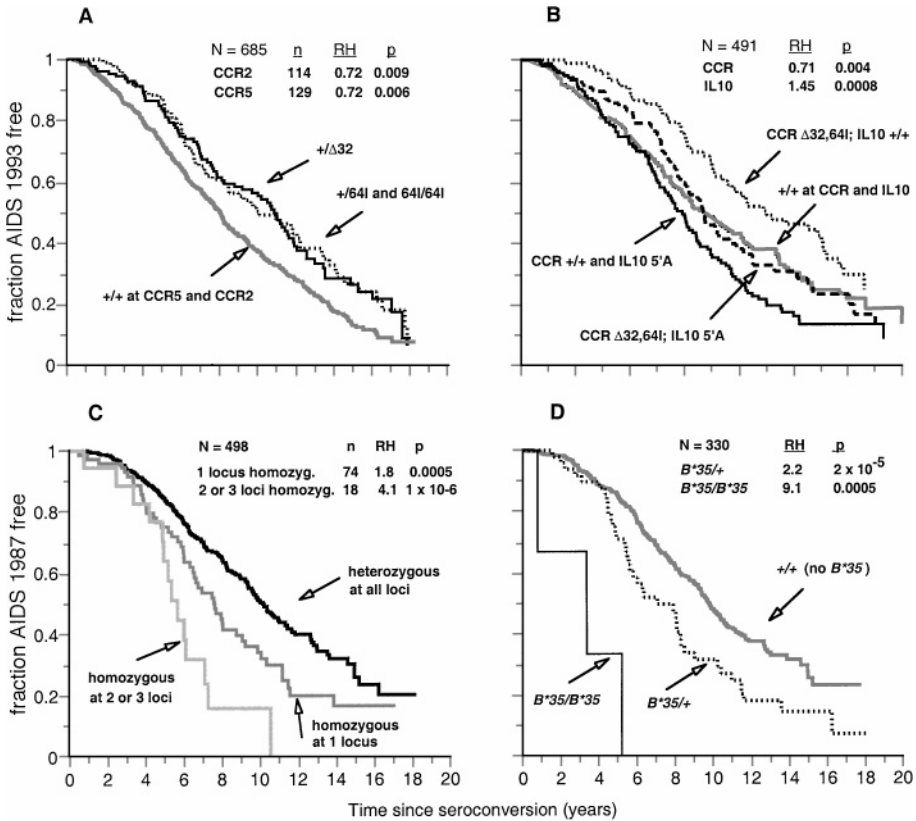
Allele	Mode	Effect	Time	Citation
1. <i>CCR5 delta 32</i>	Recessive	Prevent infection	—	24, 66, 96, 128
<i>CCR5 delta 32</i>	Dominant	Prevent lymphoma	Late	25
<i>CCR5 delta 32</i>	Dominant	Delay AIDS	Overall	24
2. <i>CCR5 P1</i>	Recessive	Accelerate AIDS	Early	70
3. <i>CCR2 641</i>	Dominant	Delay AIDS	Overall	106
4. <i>SDF1 3'A</i>	Recessive	Delay AIDS	Early	122
5. <i>HLA A, B, C</i> , "Homozygosity"	Co-dominant	Accelerate AIDS	Overall	15
6. <i>HLA B*35</i>	Co-dominant	Accelerate AIDS	Overall	15
7. <i>HLA C*04</i>	Co-dominant	Accelerate AIDS	Overall	15
8. <i>IL10 5'A</i>	Dominant	Limit Infection	—	102
<i>IL10 5'A</i>	Dominant	Accelerate AIDS	Late	102

## SURVIVAL ANALYSIS FOR DYNAMIC DISEASE PHENOTYPES

The AIDS cohorts consist of patients belonging to specific risk groups likely to be exposed to AIDS, followed clinically over the course of HIV-1 infection (12, 29, 41, 42, 45, 53, 60, 88, 116). As such the cohort populations provide a dynamic epidemiological view of the course of AIDS pathogenesis and its associated variation (24, 25, 70, 102, 106, 122). Our study group was largely Caucasian American (80%) with fewer African American, Hispanic, and Asian individuals. The median time to AIDS-defining conditions among HIV-1-infected patients was 10 years, but some people progressed to AIDS in less than a year while others avoided clinical AIDS for up to 20 years.

As the AIDS restriction genes listed in Table 2 were identified, it became evident that the rate of progression to AIDS was mediated, at least partially, by collaborating host factors; and that genetic variants in these factors could alter the kinetics of AIDS progression (hypothesis 2 in Table 1). To implicate these variants, we used a genetic modification of epidemiologic survival analysis (2, 23). First, we identified seroconverter patients, those whose date of HIV-1 infection could be estimated precisely because they had enrolled in the cohort study before converting from HIV-1 antibody-negative to antibody-positive. Their infection date was estimated as the midpoint between the last HIV-1 antibody-negative visit date and the first HIV-1 antibody-positive visit. Seroconverter patients were used to synchronize a Kaplan Meier survival curve that compared the rate of progression to AIDS "endpoints" between HIV-1-infected individuals with alternative genotypes (Figure 1). Four AIDS "endpoint" definitions were generally tracked: (a) time from HIV-1 infection to  $CD4 < 200$  cells/mm<sup>3</sup>; (b) AIDS-1993, as defined by the U.S. Centers for Disease Control (115) (that is, HIV-1 infection plus AIDS-defining illness or decline of CD4 T lymphocytes to  $< 200$  cells/mm<sup>3</sup>); (c) the more stringent AIDS-1987 definition (114) (HIV-1 infection plus AIDS-defining illness); and (d) death during follow-up for an HIV-1-infected patient. Differences between the rate of progression of different genotypes were evaluated using the Cox proportionate hazards model, which allows computation of relative hazards (RH) between genotypes, associated p-value of statistical significance, 95% confidence intervals, and cofactor interactions. (RH = 1.0 indicates no difference between genotypes; RH < 1 means the genotype slows AIDS progression; RH > 1 indicates that the genotype is associated with more rapid progression to AIDS).

All eight genes listed in Table 2 encode common natural allelic variants that exerted either a delay of AIDS (*CCR5-Δ32*, *CCR5-64I*, *SDF1-3'A*) or an acceleration to AIDS-defining conditions (*CCR5P1*, *HLA* homozygosity, *HLA-B\*35*, *HLA-Cw\*04*, *IL10-5'A*) among HIV-1-infected carriers. A survival analysis of two restriction alleles for chemokine/HIV-1 coreceptor genes (*CCR5-Δ32* and *CCR2-64I*) is illustrated in Figure 1a. Both alleles exert a dominant delay that postpones AIDS onset for 2 to 5 years. The effects are statistically significant, independent



**Figure 1** (A) Survival curves for HIV-infected persons of various *CCR5/CCR2* genotypes display the near-identical strengths of heterozygotes for *CCR5-Δ32* and *CCR2-64I* in slowing progression to AIDS (106). The data are for 685 subjects enrolled in AIDS-study cohorts whose date of seroconversion could be estimated precisely. AIDS is defined by a CDC-based 1993 standard requiring HIV infection, AIDS-defining illness, and either death or a CD4 cell count not higher than 200 (115). Each mutation delays AIDS by a mean of two or four years (compared with HIV-1-infected patients who lack both mutations). *CCR5-Δ32* occurs only in Caucasians, at about a 10% allele frequency, whereas *CCR2-64I* is ethnically ubiquitous, at allele frequencies of about 10% to 15%. (B) Survival curves demonstrating the interactive influence of dominant susceptible *IL10-5'A* bearing genotypes and *CCR2-64I* or *CCR5-Δ32* bearing genotypes on the time to AIDS-1993 in combined Caucasian cohorts (102). Numbers of patients, p-value (p), and relative hazard (RH) based on the Cox proportional hazards model are given. Cox models are based on combined analysis that considered *IL10-5'A* and *CCR5-Δ32/CCR2-64I* effects together. RH and p-values represent analyses where *CCR5/2* protective genotypes are assessed in a Cox model with *IL10* genotypes treated as covariates and vice versa. (C) Survival curves relating *HLA* class I gene homozygosity with the time from seroconversion to AIDS, using the CDC 1987 definition for AIDS (15, 114). (D) Survival curves assessing the influence of *HLA-B\*35/+* heterozygotes and *HLA-B\*35/B\*35* homozygotes in relative time to AIDS-1987 (114) compared with HIV-1-infected individuals who do not have *HLA-B\*35* alleles (15).

of each other (but see below for the functional interpretation) and cumulative in the combined and separate cohort analyses (106). The *IL10-5'A* allele exerts a dominant influence (102) whereby patients carrying the *IL10-+/5'A* and *IL10-5'A/5'A* genotypes progress to AIDS 3 to 5 years earlier than *IL10-+/+* individuals, and the effect is enhanced in the later stages (5 to 15 years post infection) (Figure 1b). The gradual development of AIDS over time actually allows a composite view of the interaction of the three genes whereby patients with *CCR2/5* protective alleles and the *IL10-+/+* genotype (i.e. *CCR5-Δ32/+*, *CCR2-64I/+*, and *IL10-+/+*) postpone AIDS the most, whereas patients with alternative susceptible genotypes progress the fastest. Patients with combinations of protective and susceptible alleles for the three loci progress at an intermediate (but still significantly different from the multilocus-protected or susceptible genotypes) rate (Figure 1b).

A dramatic influence of *HLA* class I alleles became evident in an analysis of class I gene homozygosity and specific class I alleles on survival (Figure 1c,d) (15). Among a combined cohort group of 498 seroconverters, heterozygosity itself was highly advantageous since homozygosity for any allele at either *HLA-A*, *-B*, or *-C* was highly associated with more rapid progression to AIDS (Figure 1c). Indeed, if more than one class I locus was homozygous, patients progressed to AIDS even more rapidly, with 100% succumbing by 10 years post infection. (Only half of the fully heterozygous people develop AIDS in the first 10 years.) Because nearly 75% of the rapid progressors (those who develop AIDS within four years of HIV-1 infection) are fully heterozygous at *HLA-A*, *-B*, and *-C*, we tested 63 different *HLA* alleles individually in survival analyses under recessive and dominant models for an influence on AIDS. Two alleles, *HLA-B\*35* and *Cw\*04*, both showed a codominant influence on rapid susceptibility to AIDS (Figure 1d). Thus patients with either *B\*35* or *Cw\*04* heterozygous with other alleles progress to AIDS more rapidly than other allele combinations, while homozygotes for *Cw\*04* or *B\*35* progressed the fastest, frequently within 2 to 4 years of infection.

The interactive effects of multiple loci on survival patterns would conceivably confound the interpretation of a one-locus SNP effect in analyses of patient cohorts also variable for polymorphic alleles at other restriction genes. This influence can be accounted for by two separate but equally valid approaches. First, it is possible to consider the genotypes of one locus (e.g. *CCR5*) as covariates in a survival analysis of other loci (e.g. *CCR2* or *HLA*). This option is accessible in statistical analysis software (2) designed to quantify RH estimates in the context of other covariables that exert measurable risk influence in the study population. We have utilized these weightings in recent analyses of *IL10*, *HLA*, and *CCR5P1* where previously affirmed genetic restriction of *CCR2*, *CCR5*, and *SDF1* were considered as covariates (15, 70, 102). A second approach involves simply partitioning the population to eliminate known risk-modifying genotypes. For example, one can analyze *CCR2-64I* or *SDF-3'A* genotypes in a cohort, but eliminate patients with the *CCR5-+/Δ32* protective genotypes from the analysis. In the special case of *CCR2-64I* and *CCR5-Δ32*, this was achieved automatically in the analysis in Figure 1a. It turns out that *CCR2* and *CCR5* are but 18 kb apart on chromosome

3p21 and exert complete “repulsion phase” linkage disequilibrium (i.e. *CCR5-Δ32* is invariably carried on a chromosome which is *CCR2+*, while *CCR2-64I* is always found on a chromosome with *CCR5+*) (106). Thus, only heterozygotes [*CCRS+*·*CCR2-G4I*]/[*CCRS-Δ32*·*CCR2+*], which comprise less than 1% of individuals can interact in a patient. These considerations of multiple loci become critical when the number of loci affecting the same phenotype (e.g. time to AIDS) becomes large (Table 2).

DISEASE CATEGORY ANALYSIS (DCA) TO DETECT GENETIC EPIDEMIOLOGIC SIGNALS

Perhaps the simplest approach for detecting genetic involvement in disease is to compare candidate gene allele and genotype frequencies and population HWE distribution between “disease categories”, subdivisions of the cohort with distinct clinical outcomes. Thus, to search for genetic influence on HIV-1 infection or transmission (hypothesis 1 in Table 1), HIV-1-infected individuals are compared with uninfected individuals likely to be exposed by virtue of their inclusion in risk groups for AIDS. When we originally compared allele frequencies of *CCR5-Δ32* in infected versus uninfected individuals, we found no significant differences, but there was a huge difference when we compared different *CCR5* genotypes (24, 66, 96, 128). The reason was that the homozygous *CCR5-Δ32/Δ32* genotype occurred at 1–2% among uninfected patients but was much lower, ≤0.04%, in HIV-1-infected individuals (Table 3). In our first study (24), there were no

TABLE 3 Occurrence of *CCR5* genotypes in European Caucasians

HIV-Positive				HIV-Negative				Citation
+/+	+/ <i>Δ32</i>	<i>Δ32/Δ32</i>	Total	+/+	+/ <i>Δ32</i>	<i>Δ32/Δ32</i>	Total	
1618	264	2	1883	508	87	17	612	24
645	78	0	723	582	114	8	704	96
368	93	0	461	348	82	16	446	48
348	58	0	406	203	51	7	261	78
78	22	0	100	26	7	2	35	35
291	73	0	364	—	—	—	—	49a
475	139	0	614	300	84	3	387	128
—	—	—	—	3008	744	46	3798	110
—	—	—	—	1046	165	8	1218	71
—	—	—	—	2153	492	23	2668	63
—	—	—	—	2099	384	39	2522	67
Total	3,823	727	2	4,551	10,273	2,221	169	12,652
%	84.0	16.0	0.02	81.2	17.5	1.34		

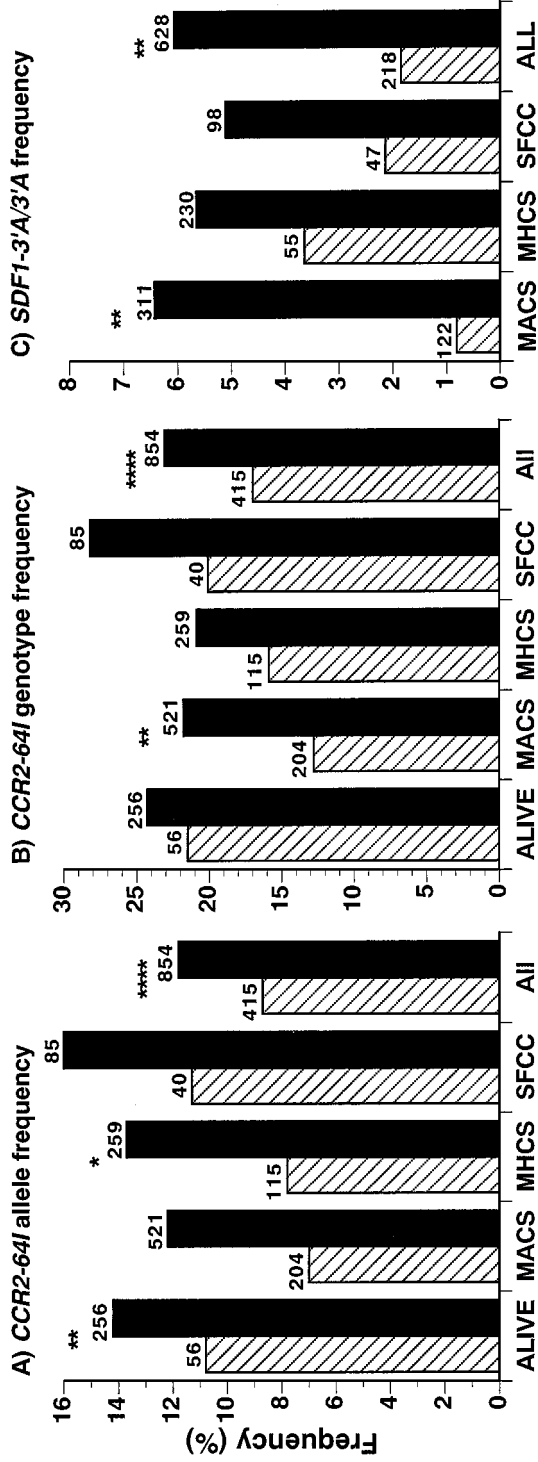
*CCR5*- $\Delta 32/\Delta 32$  among 1343 HIV-1-infected individuals, leading to a statistical p-value of  $2.5 \times 10^{-8}$ . Since that report, over 17,000 individuals have been typed for *CCR5* (Table 3) and although a few *CCR5*- $\Delta 32/\Delta 32$  homozygotes have been found to harbor HIV-1 (they became infected with a mutational derivative of HIV-1 that can utilize another chemokine receptor, CXCR4, as an entry portal instead of *CCR5*) (5, 8, 86, 101, 112), the statistical power of the distortion in genotypes between the two disease categories (infected versus uninfected) remains enormous.

The disease category analysis (DCA) can also be used to detect genetic effects within continuous phenotypic distributions such as the rate of AIDS progression (hypothesis 2 in Table 1). However, in this case the disease categories are arbitrary and defined by inspecting the survival patterns of the cohorts. Thus, one could evaluate patients' categories with different progression rates by taking the median time to AIDS for all (or each) cohorts and comparing allele/genotype/HWE in rapid progressors (who develop AIDS in a period less than the median) versus slow/non-progressors (who develop or avoid AIDS for longer than the median year interval) (Figure 2). Alternatively, we could subtract (and not consider) a few years on each side of the median, because their status is indeterminant, or we could set the cutoff time at 5 years or even 15 years.

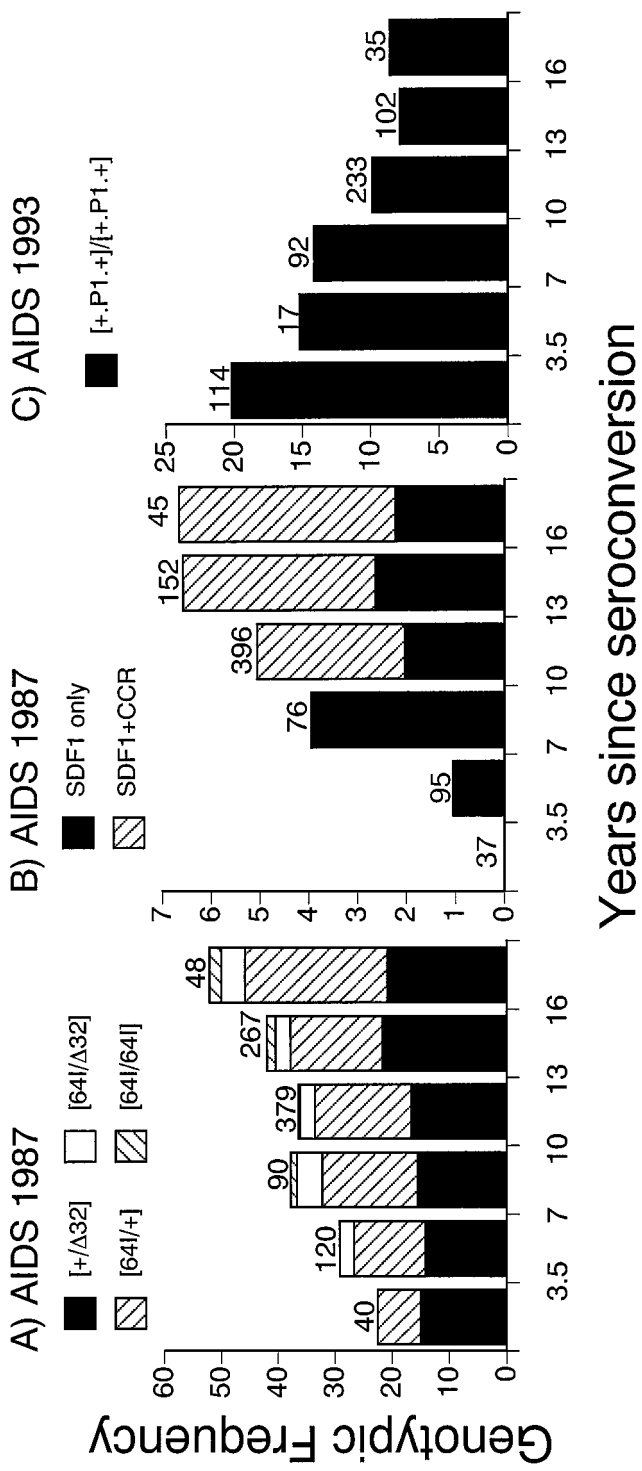
DCA uses the same patient data as the survival Kaplan Meier curves (Figure 1) and may not be as precise; yet there is an important advantage. Survival analyses must use seroconverters exclusively, yet the majority of patients enrolled in AIDS cohorts are seroprevalent; i.e. they entered the study already HIV-antibody positive. Thus we cannot estimate their date of HIV-1 infection with any confidence, and rapid progressors infected at the same time as seroprevalents would be underrepresented, since they had died before enrollment could occur (see below).

In a dichotomous DCA (one with two compared categories, e.g. rapid versus slow progressors), only seroconverters can be considered as rapid progressors since the time of infection is critical for their inclusion. However, for the slow (or non-) progressor category, we can use both seroconverters and seroprevalents. The reason is that if a seroprevalent patient has avoided AIDS until the median (say 10 years after entry into the study), then how long he was infected before entry is of no consequence; he still is a slow progressor. This consideration can increase the sample population size appreciably (in our case two- to threefold) giving remarkably increased statistical power. That aspect becomes important for independent replication in separate AIDS cohorts, a critical component for confidence in genetic associations. The confirmation effect of dichotomous DCA for *CCR2*-64I alleles, for *CCR2*-64I genotypes, and for *SDF1*-3'A genotypes in independent homosexual (MACS, SFCC), IV-drug user (ALIVE), and hemophiliac (MHCS) cohorts is illustrated in Figure 2.

A "multipoint" or "MP"-DCA can also be illustrative and useful in affirming genetic influence. In this case, the AIDS restriction genotype is measured in patients who succumb to AIDS at different times (Figure 3). The continuous distributions of AIDS development in AIDS cohorts are partitioned into discrete time intervals after HIV-1 infection: (a) AIDS < 3.5 years; (b) 3.5 to 7 years; (c) 7 to 10 years; (d) 10 to 13; (e) 13 to 16; and (f)  $\geq 16$  years. As in the dichotomous



**Figure 2** Dichotomous disease category analysis (DCA) of (A) *CCR2-64I* alleles, (B) *CCR2-+/64I* and *64I/64I* genotypes, and (C) *SDF1-3'A/3'A* genotypes. The frequency of the alleles (a) or protective genotypes (b, c) is compared between rapid progressors (solid bars) versus slow progressors (striped bars). Cutoff time is the time where half the seroconverters has progressed to AIDS-1993 (for *CCR2*) or to AIDS-1987 (for *SDF1*). Numbers of patients are listed above the bars. Slow (non) progressors include both seroconverters and seroprevalents, while rapid progressors are only seroconverter patients. Cohorts are described in (12, 29, 41, 42, 45, 53, 60, 88, 116). *CCR2* comparison (A, B) include all races; *SDF1* analysis is Caucasians only (106, 122).



**Figure 3** Multipoint disease category analysis (DCA) whereby genotype frequencies in six intervals after seroconversion are plotted and tested for statistical trends using Mantel-Haenszel  $\chi^2$  test. (A) *CCR5*- and *CCR2* protective genotypes (106). (B) *SDF1*-3'A/3'A frequency with or without *CCR2*/5 genotypic protection (122). (C) *CCR5*P1/*CCR5*P1 haplotype frequencies (70). Rapid progressors include only seroconverters; slow (non-) progressors (> 10 years to AIDS) include seroconverters and seroprevalent patients. [+P1.+] is shorthand for the [CCR2+; CCR5P1, CCR5+ (for Δ32)] haplotype.

DCA, we add both seroconverter and seroprevalent patients in the periods after 10 years. The reason is the same, that avoiding AIDS for the time since entry ( $>10$  years) is the critical point even if it is an underestimate of the actual period. In Figure 3, multipoint DCAs for four genes detected by AIDS survival analysis are presented. Figure 3a demonstrates how the frequency of both *CCR5-Δ32* and *CCR2-64I* is increased in infected study participants who avoid AIDS for increasingly longer intervals. We would predict this pattern from Figure 1a, but remember the DCA includes nearly twice as many individuals. In Figure 3b, a multipoint DCA tracks the protective homozygous *SDF1-3'A/3'A* genotype. This analysis not only confirmed the protective effect of the genotype (by showing a significant trend over time) but also makes a powerful inference about epistasis of *SDF1-3'A* with *CCR5-Δ32* and *CCR2-64I*. Although the cohorts show a median time to AIDS of 10 years for all patients irrespective of genotype, the MP-DCA indicates that every single patient with dual protection (*SDF1-3'A/3'A* plus *CCR5* or *CCR2* heterozygosity) postpones AIDS onset for at least ten years, a remarkable skew of the population distribution. In Figure 3c the multipoint DCA of the AIDS accelerating effect conferred by the 13 SNP promoter haplotype *CCR5P1/P1* shows a significantly graded decrease in the frequency in patients who avoid AIDS for longer periods. These examples illustrate how dichotomous and multipoint DCA provide powerful and confirmatory adjuncts to survival analyses by making use of heretofore untapped seroprevalent patients in association mapping.

## A FUNCTIONAL INTERPRETATION FOR AIDS RESTRICTION ALLELES

A plausible physiological context for SNP associations is critical for both credence and clinical application. For AIDS pathogenesis, an extensive empirical research effort has contributed to our understanding of each of the genes listed in Table 2, first by suggesting them as candidates and subsequently in demonstrating predicted functional differences among certain of the restriction alleles (*CCR5*, *CCR5P*, *CCR2*, *IL10*). These studies have allowed for a biologically reasonable explanation for each of the AIDS restriction alleles listed in Table 2. The role of chemokine receptors, their ligands, cytokines, and T-cell-mediated immune response on AIDS pathogenesis has been reviewed extensively (7, 19, 34, 65, 75, 84, 85). We highlight here the likely implicated mechanisms by which these allelic variants are thought to regulate the steps (Table 1) in HIV-mediated immune collapse.

In 1996, several research groups (1, 17, 18, 27, 32, 33, 37) provided cogent empirical evidence that established *CCR5* and *CXCR4* as primary coreceptors along with the T-cell recognition molecule CD4, for HIV-1 infection (Figure 4, see color insert). Between 90 and 95% of primary infections involve R5-HIV-1 strains, so-named because they interact physically with cell surface *CCR5* (after binding to

CD4), a process that triggers viral gp41-mediated virus-cell fusion (93, 95, 99, 126, 127). The virus replicates in CD4 and CCR5 bearing T-lymphocytes of multiple tissue compartments (lymph nodes, marrow, intestinal epithelium), apparently sequestered from immune clearance. Over a billion copies of virions are produced each day throughout the 10-plus year course of infection of every infected person (46, 118). Replicating viruses in most but not all patients undergo a mutational transition in the virus envelope protein, gp120, which changes the virus cell tropism from R5- to X4-HIV-1, a strain that now utilizes CD4 plus CXCR4 (instead of CCR5) to enter CD4 and CXCR4-positive T-cell subsets (93, 95, 99, 126, 127) (Figure 4). This emergence of X4-HIV-1 virus usually precedes an abrupt decline in the CD4-bearing T-lymphocytes, the hallmark of AIDS (7, 19, 65).

The *CCR5*- $\Delta$ 32 mutation encodes a 32-base pair deletion that shifts the reading frame of the *CCR5* gene coding exon leading to a premature stop codon 24 amino acid residues downstream (24, 66, 96, 128). The truncated gene product does not appear on the cell surface in *CCR5*- $\Delta$ 32/ $\Delta$ 32 homozygotes, removing the necessary entry port for the primary R5-HIV-1 strains and conferring near-complete resistance to HIV-1 infection (6, 124). The handful of exceptions (*CCR5*- $\Delta$ 32/ $\Delta$ 32 homozygotes who did become infected) have been shown to involve rare primary infections with X4-HIV-1 utilizing CXCR4 instead of CCR5 (5, 8, 86, 101, 112). *CCR5*-+/ $\Delta$ 32 heterozygotes express a diminished quantity of CCR5 on their cell surface due to gene dosage, and also because peptide heterodimers of CCR5+/CCR5- $\Delta$ 32 are retarded on the endoplasmic reticulum (6, 124). Since viral replication and spread in HIV-1-infected patients drives AIDS pathogenesis, limiting that process by diminishing CCR5 receptors would explain the slowing of the AIDS pathogenesis (Figure 1a). The effect of *CCR5* in protecting against AIDS-related B-cell lymphoma is less clear, but may involve direct HIV-1 action since CCR5 is clearly expressed on B-cell surfaces, the focal cell of the non-Hodgkins lymphoma involved in AIDS patients (25).

The promoter haplotype *CCR5P1/P1* is associated with rapid progression to AIDS and is thought to involve an increased quantity of CCR5 expression in response to tissue-specific transcription factors (14, 43, 55, 70, 73, 82). That mechanism has not been directly proven; in fact, constitutive *CCR5* expression is indistinguishable between different *CCR5P* haplotype-allele genotypes (70). However, an empirical differential of promoter allele DNA sequence binding to cREL transcription family members lends indirect support to such a scenario (10).

In addition to CCR5 and CXCR4, over a dozen chemokine receptors have been implicated as functional, albeit inefficient, coreceptors to certain rare HIV isolates (27, 32, 34, 75, 91). The *CCR2*-64I variant involves one of these minor coreceptors, and the mutation specifies a chemically conservative valine to isoleucine substitution nested within the first of seven lipophilic transmembrane regions of the molecule. In vitro studies revealed little difference between T-cells from individuals bearing the *CCR2*-+ versus *CCR2*-64I alleles in (a) the quantity of CCR2

expressed, (b) the kinetics of HIV-1 infection, or (c) specific chemokine ligand signaling of CCR2 (61, 69). A possible interpretation for the epidemiologic effect on AIDS (Figure 1a) is that alternative *CCR2* allele products interact differentially with either CCR5 or CXCR4 peptides during assembly and expression, perhaps altering their availability. An interesting but unconfirmed report (76) suggested that the *CCR2-64I* specified protein product can dimerize preferentially with CXCR4, sequestering it on the endoplasmic reticulum, whereas the wild-type *CCR2*+ specified peptides do not.

SDF-1 is a powerful chemokine produced by stromal cells, mesothelial cells, and endothelial cells that is the only identified natural ligand of CXCR4 (Figure 4). The *SDF1-3'A* variant is an SNP located in the 3' untranslated region of one of two alternative splicing transcripts SDF-1 $\beta$ . The variant is located 37 base pairs from a highly conserved sequence stretch (88–90% homologous between human and mouse), a signal for a regulatory region constrained functionally from evolutionary divergence (103, 122). A limited tissue distribution of SDF-1 in vivo and its tendency to aggregate in serum has made quantitative analysis of different genotypes difficult (4, 9, 92). We have speculated, however, that since SDF-1 is the primary ligand of CXCR4, *SDF1-3'A* may restrict the evolutionary emergence of X4-HIV-1 in situ and the ensuing AIDS accelerating process perhaps by overproduction of SDF-1 in local compartments, blocking the requisite CXCR4 receptors (122).

IL10, produced by lymphoid cells, is a powerful TH-2 cytokine that inhibits macrophage and T-cell replication, cytokine secretion from T helper cells, and HIV-1 production in macrophages (38, 39, 54, 81, 97). The *IL10-5'A* promoter variant reduces *IL10* transcription two- to fourfold and specifies a DNA sequence that fails to bind to certain ETS family transcription factors that recognize the wild-type *IL10-5'A* allele (102). *IL10-5'A*-mediated downregulation of IL10 production would release viral inhibition, predicting an increase in HIV-1 replication and more rapid progression to AIDS, which is precisely the observed epidemiologic consequence (Figure 1b).

The *HLA* homozygosity effect is predisposing to rapid AIDS onset (Figure 1c), providing a powerful affirmation of the critical influence of functional diversity at the major histocompatibility complex (MHC) in arming the cytotoxic T-lymphocytes against genetically diverse invading parasites (15, 49, 87). The concept that maximal MHC diversity increases the immune repertoire of populations and individuals has long been postulated as a driving selective force behind the high levels of MHC allelic diversity, particularly in amino acid alterations in the functional domains (i.e. peptide binding site) of MHC molecules. With a dynamic infection like HIV-1 where a swarm of mutationally diverse virions evolve to "escape" immune surveillance in every patient (83, 89), MHC diversity matters. The *HLA B\*35, -Cw\*04* acceleration is less easy to explain, since both these alleles encode class I molecules that recognize common HIV-1 amino acid signature residues. An alluring and testable hypothesis would be that these alleles provide a target for an HIV-1-encoded decoy epitope (15). In one

scenario, the host CTL machinery is diverted to defending against B\*35/C\*04 epitopes, which are effectively dispensable to the virus mission for replication and pathogenesis.

## EPIDEMIOLOGIC IMPACT OF AIDS RESTRICTION GENES

Three interactive parameters contribute to the severity of the epidemiologic consequence that each or combinations of these restriction genes would have on a study population. These include the genetic mode (i.e. is restriction dominant, recessive, codominant, or epistatic?), the population frequency of alleles and effective genotypes, and the actual strength of the effect on disease outcome. The genetic mode is determined empirically by testing all possible permutations (e.g. see Figure 1, 2). The population or cohort allele/genotype frequencies are measured directly but can be very different among ethnic groups. For example, *CCR5*- $\Delta$ 32 allele frequency ranges from 5% to 15% across different parts of Europe, but it is absent in native Africans or native East Asians. Large ethnic differences were also seen for *CCR5P*, *SDF1*, *HLA*, and *IL10* allele frequencies (15, 70, 85, 102, 122). The epidemiologic strength of the protection/susceptibility is estimated by relative hazard in the survival analysis (Figure 1) or by relative risk (RR) or odds ratio (OR) in DCA (Figure 2).

The Attributable Risk (AR) and Protected Fraction (PF) statistics were developed to assess the combinational influence of a risk factor (in our case of a particular genotype) on the entire study population (62). AR combines the three variables described above and addresses questions such as: What fraction of the slow progressors are in this category due to their carrying *CCR5*+/ $\Delta$ 32; or what fraction of the rapid progressors do so because they lack *CCR5*- $\Delta$ 32 protection? There are as many ways to formulate these questions, as there are for setting categories for DCA. Also, when more than one genotype is considered, and their effects go in opposite directions (e.g. Figure 1*b*), the computation becomes more complex.

In Table 4, we present one attempt to estimate AR for each of the eight genes on AIDS progression. The values are not additive because they interact, and mathematical models and simulations to precisely quantify the effects are under development. Statements made in the past (for example "...extended survival of 28–40% of HIV-1-infected individuals who avoid AIDS for ten or more years can be attributed to *HLA* genotype" or "...survival of 28–29% of long-term survivors, who avoid AIDS for 16 years or more are in that group because they carry *CCR5*+/ $\Delta$  32 or *CCR2*+/*64I*"...are accurate (15, 106), but require additional interactive analysis to achieve composite AR estimates and confidence limits. However, nearly 50% of the epidemiological variance in the rate of progression to AIDS can probably be attributed to the genotype for the eight restriction genes (Table 2) in the study participants (80).

**TABLE 4** Relative Hazard (RH), Relative Risk (RR), and Attributable Fractions (AF) for progression to AIDS\*

Gene	Model	Protective genotypes	Frequency	RH	RR	AF for survival past 12 years
<b>Genetic restriction</b>						
CCR5	dominant	[+/ $\Delta$ 32]	0.20	0.72	1.38	7.0%
CCR2	dominant	[+64I, 64I/64I]	0.17	0.72	1.56	8.7%
SDF1	recessive	[3'A/3'A]	0.05	0.65	1.49	2.2%
Any protective factor			0.36	0.69	1.72	20.4%
<b>Genetic susceptibility</b>						
<b>Gene</b>						
CCR5 promoter	recessive	[P1/P1]	0.13	1.52	2.02	11.6%
IL 10	dominant	[+/ $5'A$ , $5'A/5'A$ ]	0.41	1.44	1.28	10.2%
HLA Class I genes	zygosity	homozygotes at A, B, or C	0.25	1.84	1.52	11.4%
HLA-B	co-dominant	[+/ $B^*35$ , $B^*35/B^*35$ ]	0.17	1.80	1.21	3.4%
HLA-C	co-dominant	[+ $Cw^*4$ , $Cw^*04/C2^*04$ ]	0.20	1.87	1.36	6.9%
Any susceptible factor			0.60	1.72	1.61	26.7%

\* (CDC-1993 definition, Reference 115) within 5 years of HIV-1 infection or AIDS-free survival for 12 or more years, by identified AIDS restriction genes

## THE STRENGTH AND LIMITS OF ASSOCIATION ANALYSIS

The major strength of association analysis is that it uses distortions in compared population genetic frequency distribution to detect disease genes. Yet this advantage is accompanied by potential pitfalls that can lead to false positive associations as well as to missing important loci. False positive associations can arise from sampling errors in sub-structured study populations, from variation associated with multiple statistical tests, and from linkage disequilibrium (LD) of marker SNPs with the actual disease-affecting SNPs. Previous successes and increased familiarity with family-based studies to locate genes has led some to conclude that pedigree analysis is superior to, or even irreplaceable, by association analysis. However, pedigree analysis also depends on statistical departures from random association and can also lead to similar mistakes. Timing may be important in forming these impressions, particularly since monogenic disease gene discovery and theory are clearly more advanced than association mapping. Nevertheless, recent association mapping examples, including those reviewed here, lend cautious hope to the technique in implicating disease genes, by controlling for statistical or

LD artefacts. It may be relevant to mention that some of the earliest successes in disease gene inference (sickle cell anemia, Tay Sachs disease, thalassemia) were based largely on association data.

Criteria for robust validation of disease gene associations in survival, case control, or DCA are well described, but sometimes difficult to achieve (3). Optimum studies include: (a) large numbers of clinical cases, (b) replication in different study populations, (c) functional data relating the SNP association with gene function, (d) negative control SNPs in the region that fail association tests to exclude LD as an explanation, (e) low statistical p-values after conservative correction for multiple tests, (f) high relative hazards or relative risks, and (g) high attributable risk. The AIDS restriction gene studies illustrate each of these criteria in varying degrees. Our study population includes five independent cohort studies in three distinct risk groups: homosexual men (MACS, SFCC); hemophiliacs (MHCS, HGDS); and IV drug users (ALIVE). In most of our studies 2000 to 4000 patients were genotyped (15, 24, 25, 70, 102, 106, 122). Replication of epidemiologic signals among these cohorts plus in other cohort studies have been used to affirm the signals for nearly all the genes in Table 2. In addition to replicate cohorts, the AIDS genes show signals for different AIDS endpoints for survival, for dichotomous and multipoint DCA, and for Caucasian and African American ethnic groups studied separately. The p-values after corrections vary, but consistent p-values of  $10^{-2}$ – $10^{-8}$  lend confidence to the association analysis. Lastly, a functional context has been suggested for each of the AIDS restriction genes (see above), and all except *SDF1* have quantified physiological distinctiveness between wild-type and variant allele products that are consistent with the suggested function (34, 75, 85).

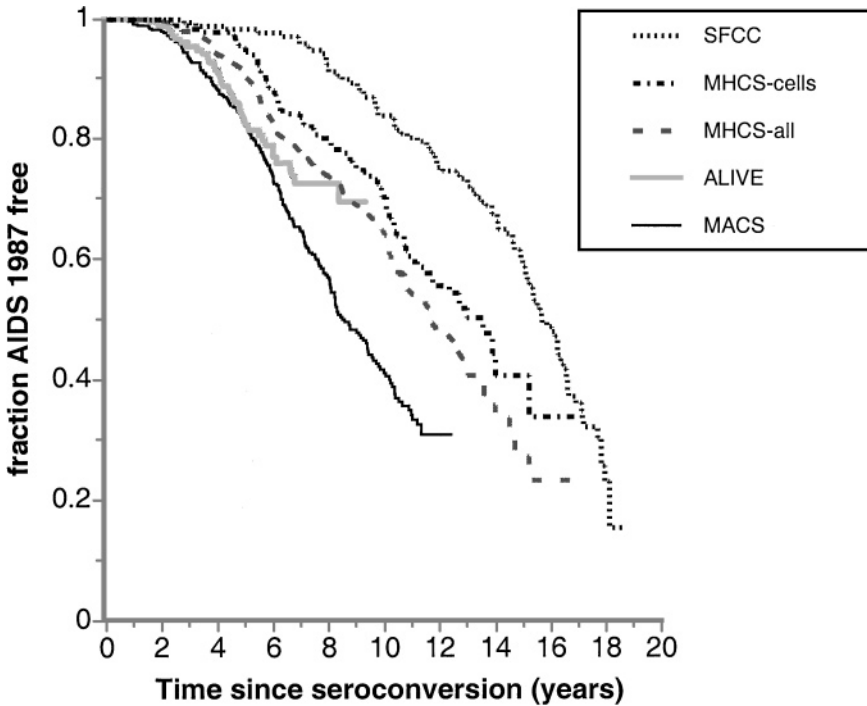
Linkage disequilibrium (the tendency of alleles at tightly linked polymorphic loci to assort nonrandomly in the populations) can lead to associations with SNPs linked to but not physiologically responsible for the clinical difference. Workers in genetic epidemiology are learning to recognize such effects and to interpret them correctly. Some examples with the AIDS cohort studies illustrate this point. First, survival analyses of nine polymorphic microsatellite loci alleles within a 10-megabase segment encompassing the *HLA* region did not reveal any homozygosity or specific allele/phenotype association with AIDS progression (15). This negative result would lend credence to the inference that the *HLA* class I epidemiological signal (Figure 1c,d) derives from the *HLA* gene products themselves and not from one of the other 221 loci within the 3.6-megabase *HLA* region (15, 77). Second, LD was extremely useful in resolving the genetic effects of *CCR5*- $\Delta$ 32, *CCR2*-64I, and *CCR5P1*, alleles of three tightly linked restriction genes located on 3p21 (14, 70, 106). These alleles occur in strong LD such that only three haplotypes including them ([*CCR2*. *CCR5P*. *CCR5*]: [64I. P1. +], [+ . P1.  $\Delta$ 32] and [+ . P1. +]) were encountered, simplifying the survival and DCA analyses considerably. That unusual LD association allowed us to exclude the postulated explanation of *CCR2*-64I restriction as arising from *CCR2*-64I LD with a down-regulating *CCR5* promoter. The reason is that the promoter allele haplotype (*CCR5P1*) in absolute LD with AIDS delaying *CCR2*-64I accelerate, rather than delay, AIDS (Figure 3c). Third, haplotype and specific SNP analysis of multiple markers in

the *CCR5P* promoter region and in the *IL10* promoter region were highly informative in implicating the responsible functionally relevant SNP (70, 102). In sum, attention to all these factors to validate association signals not only leads to confident implication of responsible SNPs, but also can be of considerable utility in excluding false positive signals.

It is also exceedingly important to be cognizant of sampling and fluctuation biases. These can be simple avoidable mistakes such as failure to correct statistically for multiple association tests (47, 68, 100, 104, 119) or combining ethnic groups with different allele frequencies of disease gene SNPs or flanking markers in LD with these. Ethnic admixture will lead to spurious distortions of population genetic parameters as a consequence of the frequency differences between the parent populations (11, 16, 109). A more subtle bias specific to time-sensitive survival analysis such as for AIDS is the "frailty" or "survival" bias that arises in cohorts established years after patients became infected or exposed (31, 105, 107). In this case, recruitment of study participants several years after HIV-1 infection would unintentionally exclude the very rapid AIDS progressors since many would have died in the interval (24, 41, 42, 45). Bias also occurs in seroconverters whose cells were collected years after infection since rapid progressors tend to have fewer viable cryopreserved lymphoid cells due to immune dysfunction. Survival or frailty effects are typified by plotting survival of seroconverters in hemophiliac cohorts (Figure 5) who show apparent (but inflated) longer survival than the gay men or IV drug user seroconverter cohorts. The reason is that the hemophiliac cohorts were established in the late 1980s, several years after introduction of HIV-1 surveillance in clotting factor lots in 1984. This bias has led to reports of associations (or absence) that, we suspect, were a consequence of the bias (51, 79).

## GENOME SCANS IN ASSOCIATION ANALYSIS

Genome-wide scans using dense maps of human SNPs and STRs have been recently suggested as a means to identify multiple gene effects of complex disease (20, 58, 94). The approach is promising because, if successful, it uncovers genes that would not be identified as candidates by other means. Very dense maps of 5000-plus STR (short tandem repeat, also termed microsatellite) loci and over 100,000 SNPs are already available, plus array technologies for high-throughput genotyping are improving daily (13, 26, 30, 44, 120). As in the candidate gene approach, the association genome scan would look for differences in population genetic parameters (allele, genotype frequency, HWE, and haplotype disposition) between unrelated disease cases and controls. Unlike linkage studies, disease gene detection by association is completely dependent upon LD between the test marker and the disease gene. That caveat is an important unknown quantity. The persistence of LD depends on the natural history of the disease gene containing chromosome segment of the test population plus the recombination distance between the disease gene and the test marker, two parameters difficult to estimate in study populations until after the disease gene is already discovered (20, 94).



**Figure 5** Survival curves of four AIDS cohorts using patients of all genotypes illustrating skews created by frailty or survival bias. ALIVE and MACS include all seroconverter patients who entered the study as HIV-1-antibody negative and converted to HIV-1-antibody positive. The SFCC was enriched intentionally for long-term nonprogressors. The MHCS is a hemophiliac cohort established in the late 1980s 3 to 5 years after seroconversion of most patients. Successfully transformed B-cell lines labeled MHCS-cells utilized to implicate AIDS restriction genes are shown separately from the entire cohort to illustrate that cell transformation is less successful in more rapid AIDS progressors (see text).

Theoretical and simulation studies can paint a bleak picture for the strength of LD, at least in homogeneous expanding human populations. Kruglyak (56) has estimated that LD is unlikely to extend beyond 3 kb in typical human populations, meaning that an efficacious genome scan would require 500,000 SNPs. These computations should be considered as over-conservative since they represent the worst-case scenario. Human genomes contain much longer stretches of LD due to histories of population bottlenecks, ethnic admixture, and selective sweeps across chromosomal regions. In our recent study of 56 Caucasian individuals in the CEPH families typed for 5048 STR loci, LD was high in many genomic regions; in all, 4% of all STR pairs within 4 cM or less displayed detectable LD (50). In the same study, a genome-wide scan for pairwise LD across 22 autosomes revealed nine chromosomal regions (2.4–6.4 cM in length) with very high levels of multilocus LD between included STR markers in a 5 cM sliding window, likely persisting

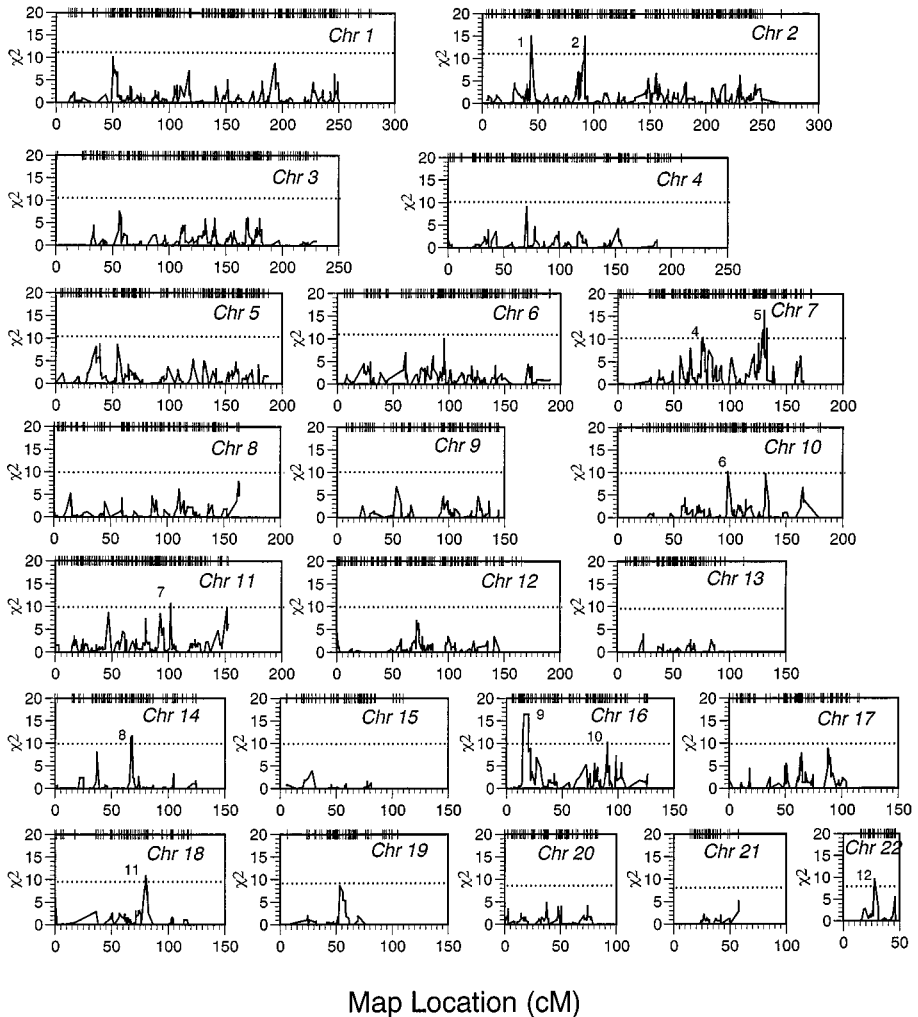
from recent selective sweeps in the history of Caucasian population (Figure 6). Subsequent theoretical treatments and empirical approaches indicate that as few as 30,000 markers (at an average density of 100 kb) would suffice for LD association mapping (20).

Our own studies of flanking markers of the AIDS restriction genes have detected significant LD and association signals using adjacent markers spanning 4000 bp for *IL10* and 1 cM for *CCR5*, *CCR2*, and *CCR5P1* (14, 70, 102, 106, 110). The *IL10-5'A* influence was originally discovered by association analysis of 19 STRs linked (1–1000 kb) to a group of candidate loci (102). Among 17 candidate genes tested, two STRs linked to *CCR5* and one linked to *IL10* were the only ones to show a signal. When these 19 STRs are added to the 25 STRs that showed no association in the *HLA-AIDS* analysis (15), 3 STRs of 44 showed an epidemiological signal for AIDS progression. Two were derived from LD with *CCRS-Δ32* and the third from *IL10-5'A*, functional AIDS restriction alleles (Table 2).

It is possible to increase the power of association mapping by selecting populations (and markers) that would be expected to show increased LD (123). One option is to study recently founded populations who have passed through a narrow population bottleneck placing much of the genome, particularly markers with rare alleles, into LD. Such genetic “drift mapping” has real potential in increasing the resolving power of association genome scans, because the founder effect increases the length of regions in LD appreciably (57, 111), in some cases up to 10 cM. Several well-known examples of founded small populations (Finland, Iceland, Sardinia, Japan, and others) show these effects (123). But many other less well described populations retain extensive LD because they have descended from cyclic expansions following demographic contractions or founder events from human migrations (20, 123).

A second powerful approach is to sample recently admixed populations between different ethnic groups (11, 16, 74, 109). Admixed populations generate LD between loci whose allele frequencies are markedly different in the parent populations as a consequence of gene flow. Theoretical and simulation studies of MALD (mapping by admixture linkage disequilibrium) predict that LD segments of up to 20 cM would be retained in continually admixed populations such as African Americans or Hispanics (11, 109). Optimistic scenarios would suggest that genome scans of 10-cM spaced markers (300–600 markers) enriched for high delta values (delta is the difference in allele frequency between homologous alleles from the parent ethnic groups that contributed to the admixture) would detect 95% of the disease gene variants that themselves show a high delta value (109). A recent experimental validation of this approach revealed a 20 cM stretch of LD, including 8 STRs, around the FY (Duffy) locus in sampled African Americans (59).

MALD has been applied to diabetes in American Indians admixed with Europeans (74) and to hypertension in African Americans (125). The ALIVE AIDS cohort, which is predominately African American IV drug users, also holds considerable promise for this method (116). MALD has two possible disadvantages.



**Figure 6** Distribution of linkage disequilibrium (LD) among linked STR loci in 5-cM sliding windows across the human genome (50). The X-axis is the chromosomal location of the anchor locus (the first “upstream” locus in the window) based on the recombination linkage map; the Y-axis shows the  $\chi^2$  with 1 d.f. for the probability of the observed or greater LD within each 5.0-cM window. Tick marks above plots indicate STR locus positions. The horizontal dotted line indicates the height of the *HLA* region on chromosome 6 as a reference LD threshold to identify ten other regions of high interlocus LD.

First, the delta value of the disease gene variants must be high ( $\delta \geq 0.3$ ) to be detectable and this may not apply for all disease genes. Second, the marker signal may actually reveal an LD segment that is rather large (e.g. 5–20 cM), making positional cloning or responsible gene SNP implication formidable. However, narrowing a locus to a chromosome segment is a welcome first step in haplotype analysis to implicate the responsible disease gene locus.

## CONCLUSIONS

Although population genetic theory and practice were not developed as mapping tools, their usefulness in this new field is assured. Carefully designed and executed association analysis can be nearly as sensitive as linkage studies in revealing polygenic and multifactorial phenotypes, and in some cases more so. Aspects that can influence success are sample size, population demographic history, rigorous clinical description of phenotypes, plausible biologic context for associations, appropriate statistical methodology, and adequate control for ethnic, population, or cohort perturbations unrelated to the disease.

The AIDS restriction gene discoveries have revealed both strengths and pitfalls of association analysis. The strength would include the availability of large numbers of patients, exquisite clinical histories and descriptions, widespread cooperation from the HIV-1-infected community, and elegant basic research on the AIDS pathogenic process. Connecting the human gene polymorphisms to this disease has illuminated the critical role that the host plays in the decline of immune function. Indeed, nearly half of the epidemiologic variance or heterogeneity can be attributed to the composite genotype of study participants. And yet we suggest that many more gene variants will be uncovered in future years that implicate steps 1, 3, 4, and 5 in Table 1. Then, our genes that have evolved to defend against infectious disease, their functional variants, and perhaps useful clues to therapeutic intervention of this deadly scourge will be revealed.

**Visit the Annual Reviews home page at [www.AnnualReviews.org](http://www.AnnualReviews.org)**

## LITERATURE CITED

1. Alkhatib G, Combadiere C, Broder CC, Feng Y, Kennedy PE, et al. 1996. CC CKR5: A RANTES, MIP-1 $\alpha$ , MIP-1 $\beta$  receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* 272:1955–58
2. Allison PD. 1995. *Survival Analyses Using the SAS System: A Practical Guide*, pp. 155–97. Cary, NC: SAS Inst.
3. Anonymous. 1999. Freely associating. *Nat. Genet.* 22:1–2
4. Bajetto A, Bonavia R, Barbero S, Piccioli P, Costa A, et al. 1999. Glial and neuronal cells express functional chemokine receptor CXCR4 and its natural ligand stromal cell-derived factor 1. *J. Neurochem.* 73:2348–57

5. Balotta C, Bagnarelli P, Violin M, Ridolfo AL, Zhou D, et al. 1997. Homozygous  $\Delta 32$  deletion of the *CCR-5* chemokine receptor gene in an HIV-1 infected patient. *AIDS* 11:F67-71
6. Benkirane M, Jin DY, Chun RF, Koup RA, Jeang KT. 1997. Mechanism of transdominant inhibition of CCR5-mediated HIV-1 infection by *CCR5* $\Delta 32$ . *J. Biol. Chem.* 272:30603-6
7. Berger EA, Murphy PM, Farber JM. 1999. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu. Rev. Immunol.* 17:657-700
8. Biti R, French R, Young J, Bennets B, Stewart G, et al. HIV-1 infection in an individual homozygous for the *CCR5* deletion allele. *Nat. Med.* 3:252-53
9. Bleul CC, Fuhlbrigge RC, Cassanovas JM, Aiuti A, Springer TA. 1996. A highly efficacious lymphocyte chemoattractant, stromal cell-derived factor 1 (SDF-1). *J. Exp. Med.* 184:1101-9
10. Bream JH, Young HA, Rice N, Martin MP, Carrington M, et al. 1999. *CCR5* promoter alleles distinguished by specific DNA binding factors. *Science* 284:223a
11. Briscoe D, Stephens JC, O'Brien SJ. 1994. Linkage disequilibrium in admixed populations: applications in gene mapping. *J. Hered.* 85:59-63
12. Buchbinder SP, Katz MH, Hessel NA, O'Malley PM, Holmberg SD. 1994. Long-term HIV-1 infection without immunologic progression. *AIDS* 8:1123-28
13. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, et al. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22:231-38
14. Carrington M, Dean M, Martin MP, O'Brien SJ. 1999. Genetics of HIV-1 infection: chemokine receptor *CCR5* polymorphism and its consequences. *Hum. Mol. Genet.* 8:1939-45
15. Carrington M, Nelson G, Martin MP, Kissner T, Vlahov D, et al. 1999. *HLA* and HIV-1: heterozygote advantage and *B\*35-Cw\*04* disadvantage. *Science* 283:1748-52
16. Chakraborty R, Weiss KM. 1988. Admixture as a tool for finding linked genes and detecting the difference from allelic association between loci. *Proc. Natl. Acad. Sci. USA* 85:9119-23
17. Choe H, Farzan M, Sun Y, Sullivan N, Rollins B, et al. 1996. The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* 85:1135-48
18. Cocchi F, De Vico AL, Garzino-Demo A, Arya SK, Gallo RC, et al. 1995. Identification of RANTES, MIP-1 $\alpha$ , and MIP-1 $\beta$  as the major HIV-suppressive factors produced by CD8+ T cells. *Science* 270:1811-15
19. Cohen OJ, Fauci AS. 2000. Pathogenesis and medical aspects of HIV-1 infection. In *Fields Virology*, ed. BN Fields, DM Knipe, PM Howley. New York: Lippincott, Williams & Wilkins
20. Collins A, Lonjou C, Morton NE. 1999. Genetic epidemiology of single-nucleotide polymorphisms. *Proc. Natl. Acad. Sci. USA* 96:15173-77
21. Collins FS, Guyer MS, Chakravarti A. 1997. Variations on a theme: cataloging human DNA sequence variation. *Science* 278:1580-81
22. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Waiters L, and the members of the DOE and NIH planning groups. 1998. New goals for the U.S. Human Genome Project: 1998-2003. *Science* 282:682-89
23. Cox DR, Oakes D. 1984. *Analysis of Survival Data*. New York: Chapman & Hall
24. Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, et al. 1996. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CCR5* structural gene. *Science* 273:1856-62
25. Dean M, Jacobson LP, McFarlane G,

- Margolick JB, Jenkins FJ, et al. 1999. Reduced risk of AIDS lymphoma in individuals heterozygous for the *CCR5*- $\Delta$ 32 mutation. *Cancer Res.* 59:3561-64
26. Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, et al. 1998. A physical map of 30,000 human genes. *Science* 282:744-46
27. Deng HK, Liu R, Ellmeier W, Choe S, Utumaz D, et al. 1996. Identification of a major co-receptor for primary isolates of HIV-1. *Nature* 381:661-66
28. Deleted in proof
29. Detels R, Liu Z, Hennessey K, Kan J, Visscher BR, et al. 1996. For the Multicenter AIDS Cohort Study: resistance to HIV-1 infection. *J. AIDS* 7:1263-69
30. Dib C, Faure S, Fizames C, Samson D, Drouot N, et al. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152-54
31. Donfield S, Lynn HS, Hilgartner MW. 1998. Progression to AIDS. *Science* 280:1819-20
32. Doranz BJ, Rucker J, Yi Y, Smyth RJ, Samson M, et al. 1996. A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. *Cell* 85:1149-58
33. Dragic T, Litwin V, Allaway GP, Martin SR, Huang Y, et al. 1996. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature* 381:667-73
34. D'Souza MP, Harden VA. 1996. Chemokines and HIV-1 second receptors-confluence of two fields generates optimism in AIDS research. *Nat. Med.* 2:1293-300
35. Eugenolsen J, Iverson AKN, Garred P, Koppelhus U, Pedersen C, Benfield TL, et al. 1997. Heterozygosity for a deletion in the CKR-5 gene leads to prolonged AIDS-free survival and slower CD4 T-cell decline in a cohort of HIV-1 seropositive individuals. *AIDS* 11:305-10
36. Evans WE, Relling MV. 1999. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 286:487
37. Feng Y, Broder CC, Kennedy PE, Berger EA: 1996. HIV-1 entry cofactor: Functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* 272:872-77
38. Fiorentino DF, Bond MW, Mosmann TR. 1989. Two types of mouse helper T cell. IV. Th2 clones secrete a factor that inhibits cytokine production by Th1 clones. *J. Exp. Med.* 170:2081-95
39. Fiorentino DF, Zlotnik A, Mosmann TR, Howard M, O'Garra A. 1991. IL10 inhibits cytokine production by activated macrophages. *J. Immunol.* 147:3815-22
40. Glavac D, Dean M. 1995. Applications for heteroduplex analysis for mutations detection in disease genes. *Hum. Mutat.* 6:281-87
41. Goedert JJ, Biggar RJ, Winn DM, Mann DL, Byar DP, et al. 1985. Decreased helper T lymphocytes in homosexual men. I. Sexual contact in high-incidence areas for the acquired immunodeficiency syndrome. *Am. J. Epidemiol.* 121:629-36
42. Goedert JJ, Kessler CM, Aledort LM, Biggar RJ, Andes WA, et al. 1989. A prospective study of human immunodeficiency virus type 1 infection and the development of AIDS in subjects with hemophilia. *N. Engl. J. Med.* 321:1141-48
43. Gonzalez E, Bamshad M, Sato N, Mummid S, Dhanda R, et al. 1999. Race-specific HIV-1 disease-modifying effects associated with *CCR5* haplotypes. *Proc. Natl. Acad. Sci. USA* 96:12004-9
44. Halushka MK, Fan J-B, Bently K, Hsie L, Shen N, et al. 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat. Genet.* 22:239-47
45. Hilgartner MW, Donfield SM, Willoughby A, Contant CF Jr, Evatt BL, et al. 1993. Hemophilia growth and development

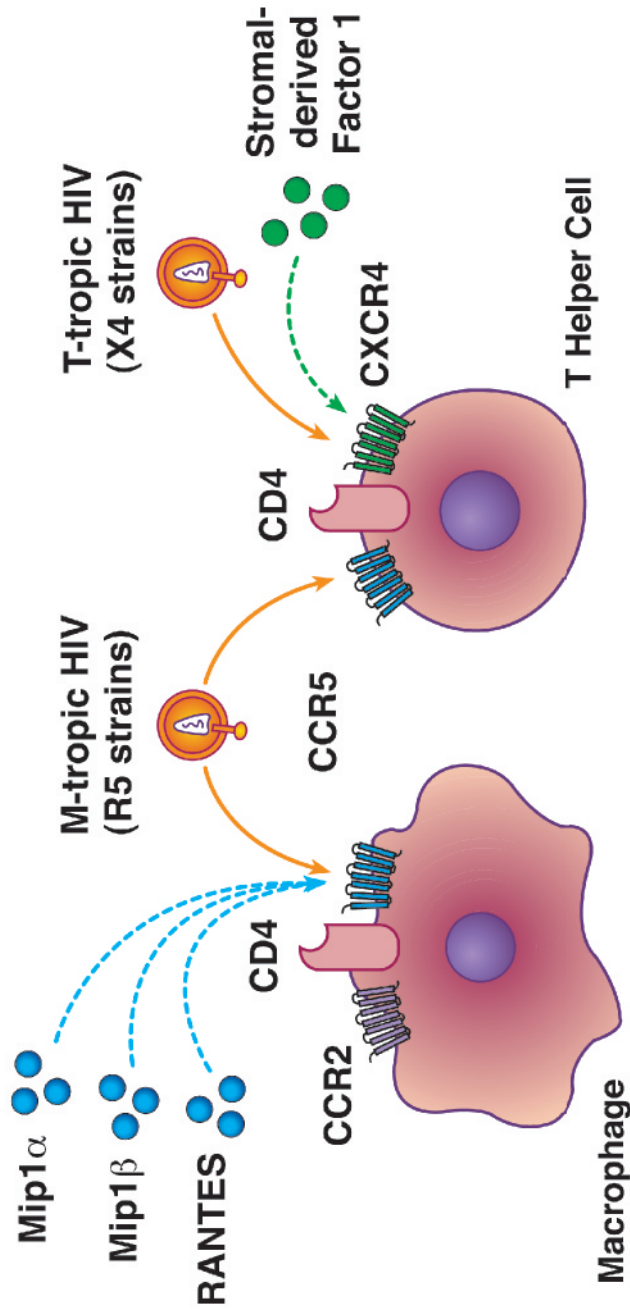
- study. Design, methods, and entry data. *Am. J. Pediatr. Hematol. Oncol.* 15:208–18
46. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, et al. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373:123–26
47. Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* 6:65–70
48. Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T, et al. 1996. The role of a mutant *CCR5* allele in HIV-1 transmission and disease progression. *Nature Med.* 2:1240–1243.
49. Hughes A, Yeager M. 1998. Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genet.* 32:415–35
- 49a. Husman AMD, Koot M, Cornelissen M, Keet JPM, Brouwer M, et al. 1997. Association between *CCR5* genotype and the clinical course of HIV-1 infection *Ann. Intern. Med.* 127:882–90
50. Huttley GA, Smith MW, Carrington M, O'Brien SJ. 1999. A scan for linkage disequilibrium across the human genome. *Genetics* 152:1711–22
51. Ioannidis JPA, Rosenberg PS, Goedert JJ, Kim R, Ahuja SK, et al. Effects of *CCR5* and *CCR2* polymorphisms on the rate of HIV-1 disease progression: An international meta-analysis. Submitted
52. Jones AC, Austin J, Hansen N, Hoogendoorn B, Oefner PJ, et al. 1999. Optimal temperature selection for mutation detection by denaturing HPLC and comparison to single-stranded conformation polymorphism and heteroduplex analysis. *Clin. Chem.* 45:1133–40
53. Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, et al. 1987. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Am. J. Epidemiol.* 126:310–18
54. Kollmann TR, Pettoello-Mantovani M, Katopodis NF, Hachamovitch M, Rubinstein A, et al. 1996. Inhibition of acute in vivo human immunodeficiency virus infection by human interleukin 10 treatment of SCID mice implanted with human fetal thymus and liver. *Proc. Natl. Acad. Sci. USA* 93:3126–31
55. Kostrikis LG, Huang Y, Moore JP, Wolinsky SM, Zhang L, et al. 1998. A chemokine receptor *CCR2* allele delays HIV-1 disease progression and is associated with a *CCR5* promoter mutation. *Nat. Med.* 4:350–53
56. Kruglyak L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* 22:139–44
57. Laan M, Paabo S. 1998. Mapping genes by drift-generated linkage disequilibrium in human populations. *Nat. Genet.* 17:435–38
58. Lander ES, Schork NJ. 1994. Genetic dissection of complex traits. *Science* 265:2037–47
59. Lautenberger JA, Stephens JC, O'Brien SJ, Smith MW. 2000. Significant admixture linkage disequilibrium across 30 cM around the FY locus in African Americans. *Am. J. Hum. Genet.* 66:969–78
60. Lederman MM, Jackson JB, Kroner BL, White GC III, Eyster ME, et al. 1995. Human immunodeficiency virus (HIV) type 1 infection status and in vitro susceptibility to HIV infection among high-risk HIV-1-seronegative hemophiliacs. *J. Infect. Dis.* 172:228–31
61. Lee B, Doranz BJ, Rana S, Yi Y, Mellado M, et al. 1998. Influence of the *CCR2-V64I* polymorphism on human immunodeficiency virus type 1 coreceptor activity and on chemokine receptor function of *CCR2b*, *CCR3*, *CCR5*, and *CXCR4*. *J. Virol.* 72:7450–58
62. Levin ML. 1953. The occurrence of lung cancer in man. *Acta Unio Int. Contra Cancrum* 9:531–41
63. Libert F, Cochaus P, Beckman G,

- Samson M, Aksenova M, Cao A, Czeizel A, et al. 1998. The  $\Delta$ CCR5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. *Hum. Mol. Genet.* 7:399–406
64. Lipshutz RJ, Fodor SP, Gieras TR, Lockhart DJ. 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* 21:20–24
65. Littman DR. 1998. Chemokine receptors: keys to AIDS pathogenesis? *Cell* 93:677–80
66. Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, et al. 1996. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 86:367–77
67. Lucotte G. 1997. Frequencies of the CC chemokine receptor 5  $\Delta$ 32 allele in various populations of defined racial background. *Biomed. Pharmacother.* 51:469–73
68. Ludbrook J. 1998. Multiple comparison procedures updated. *Clin. Exp. Pharm. Physiol.* 25:1032–37
69. Mariani R, Wong S, Mulder LCF, Wilkinson DA, Reinhart AL, et al. 1999. CCR2-64I polymorphism is not associated with altered CCR5 expression or coreceptor function. *J. Virol.* 73:2450–59
70. Martin MP, Dean M, Smith MW, Gerrard B, Michael NL, et al. 1998. Genetic acceleration of AIDS progression by a promoter variant of CCR5. *Science* 282:1907–11
71. Martinson JJ, Chapman NH, Rees DC, Liu Y-T, Clegg JB. 1997. Global distribution of the CCR5 gene 32-basepair deletion. *Nat. Genet.* 16:100–02.
72. Masood E. 1999. As consortium plans free SNP map of human genome. *Nature* 398:545–46
73. McDermott DH, Zimmerman PA, Guignard F, Kleeberger CA, Leitman SF, et al. 1998. CCR5 promoter polymorphism and HIV-1 disease progression. Multicenter AIDS Cohort Study (MACS). *Lancet* 352:866–70
74. McKeigue PM. 1998. Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. *Am. J. Hum. Genet.* 63:241–51
75. McNicholl JM, Smith DK, Qari SH, Hodge T. 1997. Host genes and HIV: the role of the chemokine receptor gene CCR5 and its allele ( $\Delta$ 32 CCR5). *Emerg. Infect. Dis.* 3:261–71
76. Mellado M, Ridriguez-Frade JM, Vila-Coro AJ, de Ana AM, Martinez-A C. 1999. Chemokine control of HIV-1 infection. *Nature* 400:723–24
77. MHC Sequencing Consortium 1999. Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401:921–23
78. Michael NL, Chang G, Louie LG, Mascola JR, Dondero D, et al. 1997. The role of viral phenotype and CCR-5 gene defects in HIV-1 transmission and disease progression. *Nat. Med.* 3:338–40
79. Michael NL, Louie LG, Rohrbaugh AL, Schultz KA, Dayhoff DE, et al. 1997. The role of CCR5 and CCR2 polymorphisms in HIV-1 transmission and disease progression. *Nat. Med.* 3:1160–62
80. Miettinen OS. 1974. Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am. J. Epidemiol.* 99:325–32
81. Muller F, Aukrust P, Lien E, Haug CJ, Froland SS. 1998. Enhanced interleukin-10 production in response to *Mycobacterium avium* products in mononuclear cells from patients with human immunodeficiency virus infection. *J. Infect. Dis.* 177:586–94
82. Mummidi S, Ahuja SS, Gonzalez E, Anderson SA, Santiago EN, et al. 1998. Genealogy of the CCR5 locus and chemokine system gene variants associated with altered rates of HIV-1 disease progression. *Nat. Med.* 4:786–93
83. Nowak MA, Anderson RM, McLean AR,

- Wolfs TFW, Goudsmit J, et al. 1991. Antigenic diversity thresholds and the development of AIDS. *Science* 254:963–69
84. O'Brien SJ. 1998. AIDS: a role for host genes. *Hosp. Pract.* 33:53–79
85. O'Brien SJ, Moore JP. 2000. The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS. *Immunol. Rev.* In press
86. O'Brien TR, Winkler C, Dean M, Nelson JAE, Carrington M, et al. 1997. HIV-1 infection in a man homozygous for *CCR5*- $\Delta 32$ . *Lancet* 349:1219
87. Parham P, Ohta T. 1996. Population biology of antigen presentation by MHC class I molecules. *Science* 272:67–74
88. Phair J, Jacobson L, Detels R, Rinaldo C, Saah A, et al. 1992. Acquired immune deficiency syndrome occurring within 5 years of infection with human immunodeficiency virus type-1: The Multicenter AIDS Cohort Study. *J. AIDS* 5:490–96
89. Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, et al. 1991. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* 354:453–59
90. Poduslo SE, Dean M, Kolch U, O'Brien SJ. 1991. Detecting high-resolution polymorphisms in human coding loci by combining PCR and single-strand conformation polymorphism (SSCP) analysis. *Am. J. Hum. Genet.* 49:106–11
91. Premack BA, Schall TJ. 1996. Chemokine receptors: gateways to inflammation and infection. *Nat. Med.* 2:1174–78
92. Rempel SA, Dudas S, Ge S, Gutierrez JA. 2000. Identification and localization of the cytokine SDF1 and its receptor, CXCR4 chemokine receptor 4, to regions of necrosis and angiogenesis in human glioblastoma. *Clin. Cancer Res.* 6:102–11
93. Richman DD, Bozzette SA. 1994. The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. *J. Infect. Dis.* 169:968–74
94. Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. *Science* 273:1516–17
95. Roos MTL, Lange JMA, de Goede REY, Coutinho RA, Schellekens PTA, et al. 1992. Viral phenotype and immune response in primary human immunodeficiency virus type 1 infection. *J. Infect. Dis.* 165:427–32
96. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, et al. 1996. Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the *CCR-5* chemokine receptor gene. *Nature* 382:722–25
97. Schols D, De Clercq E. 1996. Human immunodeficiency virus type 1 gp120 induces anergy in human peripheral blood lymphocytes by inducing interleukin-10 production. *J. Virol.* 70:4953–60
98. Schriml LM, Peterson RJ, Gerrard B, Dean M. 2000. Use of denaturing HPLC to map human and murine genes and to validate single-nucleotide polymorphisms. *BioTechniques* 28:740–45
99. Schuitemaker H, Koot M, Kootstra NA, Dergksen MW, de Goede REY, et al. 1992. Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: Progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus populations. *J. Virol.* 66:1354–60
100. Scweder TE, Spjøtvoll E. 1982. Plots of P-values to evaluate many tests simultaneously. *Biometrika* 69:493–502
101. Sheppard H, Celum C, Michael N, O'Brien SJ, Carrington M, et al. HIV-1 infection in individuals with the *CCR5*- $\Delta 32/\Delta 32$  genotype: acquisition of syncytium inducing virus at seroconversion. *J. Infect. Dis.* Submitted
102. Shin HD, Winkler C, Bream J, Young H, Phair J, et al. Genetic restriction of HIV-1 infection and AIDS progression by promoter alleles of interleukin 10. *Proc. Natl. Acad. Sci. USA.* In press

103. Shirozu M, Nakano T, Inazawa J, Tashiro K, Tada H, et al. 1995. Structure and chromosomal localization of the human stromal cell-derived factor 1 (SDF1) gene. *Genomics* 28:495–500
104. Sidak Z. 1967. Rectangular confidence regions for the means of multivariate normal distributions. *J. Am. Statist. Assoc.* 62:626–33
105. Smith MW, Carrington M, Winkler C, Lomb D, Dean M, et al. 1997. CCR2 chemokine receptor and AIDS progression. *Nat. Med.* 1997; 3:1052–53
106. Smith MW, Dean M, Carrington M, Winkler C, Huttley GA, et al. 1997. Contrasting genetic influence of CCR2 and CCR5 receptor gene variants on HIV-1 infection and disease progression. *Science* 277:959–65
107. Smith MW, Dean M, Carrington M, Winkler C, O'Brien SJ. 1998. Progression to AIDS. *Science* 280:1819–20
108. SNP—Single nucleotide polymorphism web addresses: <http://www.ncbi.nlm.nih.gov/ncicgap/><http://lpg.nci.nih.gov/GAI/><http://www.ncbi.nlm.nih.gov/SNP/><http://hgbase.interactiva.de/>
109. Stephens JC, Briscoe D, O'Brien SJ. 1994. Mapping by admixture linkage disequilibrium (MALD) in human populations: limits and guidelines. *Am. J. Hum. Genet.* 55:809–24
110. Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, et al. 1998. Dating the origin of the CCR5-Δ32 AIDS resistance gene allele by the coalescence of haplotypes. *Am. J. Hum. Genet.* 62:1507–15
111. Terwilliger JD, Zollner S, Laan M, Paabo S. 1998. Mapping genes through the use of linkage disequilibrium generated by genetic drift, “drift mapping” in small populations with no demographic expansion. *Hum. Hered.* 48:138–54
112. Theodorou I, Meyer L, Magierowska M, Katlama C, Rouzioux C. 1997. HIV-1 infection in an individual homozygous for CCR5-Δ32. *Lancet* 349:1219–20
113. Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, et al. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high performance liquid chromatography. *Genome Res.* 7:996–1005
114. US Cent. Dis. Control. 1987. *Morb. Mortal. Wkly. Rep.* 36 (Suppl. 1)
115. US Cent. Dis. Control Prevent. 1992. *Morb. Mortal. Wkly. Rep.* 41, 1
116. Vlahov D, Anthony JC, Munoz A, Margolick J, Nelson KE, et al. 1991. The ALIVE study, a longitudinal study of HIV-1 infection in intravenous drug users: description of methods and characteristics of participants. *NIDA Res. Monogr.* 109:75–100
117. Wang DG, Fan JB, Siao CJ, Berno A, Young P, et al. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077–82
118. Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, et al. 1995. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 373:117–22
119. Weir BS. 1996. *Genetic Data Analysis*. Sunderland, MA: Sinauer
120. Weissbach J. 1998. The Human Genome Project: from mapping to sequencing. *Clin. Chem. Lab. Med.* 36:511–14
121. Wickelgren I. 1999. Mining the genome for drugs. *Science* 285:998
122. Winkler C, Modi W, Smith MW, Nelson GW, Wu X, et al. 1998. Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. *Science* 279:389–93
123. Wright AF, Carothers AD, Pirastu M. 1999. Population choice in mapping genes for complex diseases. *Nat. Genet.* 23:397–404
124. Wu L, Paxton WA, Kassam N, Ruffing N, Rottmen JB, et al. 1997. CCR5 levels and expression pattern correlate with

- infectibility by macrophage-tropic HIV-1, in vitro. *J. Exp. Med.* 185:1681–91
125. Zebra KE, Ferrell RE, Sing CF. 1998. Genetic structure of five susceptibility gene regions for coronary artery disease: disequilibria within and among regions. *Hum. Genet.* 103:346–54
126. Zhang LQ, Mackenzie P, Cleland A, Holmes EC, Leigh-Brown AJ, et al. 1993. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *J. Virol.* 67:3345–56
127. Zhu T, Mo H, Wang H, Nam SD, Cao Y, et al. 1993. Genotypic and phenotypic characterization of HIV-1 in patients with primary infection. *Science* 261:1179–81
128. Zimmerman PA, Bukckler-White A, Alkhatib G, Spalding T, Kubofcik J, et al. 1997. Inherited resistance to HIV-1 conferred by an inactivating mutation in CC chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background, and quantified risk. *Mol. Med.* 3:23–36



**Figure 4** The clinical consequence of HIV-1 infection appears to depend in large part on the pathogen's changing abilities to use host cell-surface receptors as a means of entering cells. The infecting virus is almost always R5 and macrophage-tropic: It requires the host cell's expression of CD4 and CCR5, whose normal ligands are three chemokines, RANTES, SDF1, MIP-1 $\alpha$ , and MIP-1 $\beta$ . In most patients studied, a mutational alteration in the viral envelope, gp120, occurs and leads to an alteration in coreceptor preference so that CXCR4 is used instead of CCR5. This late-stage X4-tropic HIV-1 usually appears just prior to the collapse of the CD4 bearing T-lymphocyte subset. The natural ligand of the CD4 bearing T-lymphocyte subset is a stromal disease factor.